Chapter 12
Architectures and Information Signaling Techniques for Cognitive Networks

Dzmitry Kliazovich
University of Luxembourg, Luxembourg

Nelson Fonseca
State University of Campinas, Brazil

Fabrizio Granelli
University of Trento, Italy

Pascal Bouvry
University of Luxembourg, Luxembourg

ABSTRACT

The introduction of self-awareness, self-management, and self-healing into networks leads to a novel paradigm known as cognitive networking. This chapter overviews recent developments of this concept, discussing possible cognitive node structure and candidate cognitive network architecture. It explains the functionality of cognitive algorithms and discusses opportunities for potential optimization. Furthermore, the concept of cognitive information services is introduced. Information signaling techniques are then classified, reviewed in details, and compared among them. Finally, the performance of cognitive communication protocols is presented for a choice of examples.

INTRODUCTION

The evolution of communication technologies has brought networking a step closer to a service provisioning on an "anytime, anywhere" basis, yet ensuring instantaneous secure communications. However, such an innovation is limited by the constraints imposed by the original design of the Internet which leads to the inefficient configuration and management of networks (Georgakopoulos, Tsagkaris, Karvounas, Vlacheas, & Demestichas, 2012).

Self-awareness, self-management, and self-healing have all been proposed for the optimization of network operation, reconfiguration, and

management. The introduction of such "self-properties" contributes to an increase in network "intelligence", as well as creating a new paradigm, known as cognitive networking, which is expected to be part of the 4th generation wireless networks (4G) (Maravedis, 2006).

Cognition implies the awareness of the network of its own operational state, as well as its ability to automatically adjust operational parameters to accomplish specific tasks, such as the detection of environmental changes. Cognition depends on the support of network elements (routers, switches, base stations, etc.), to host active agents which can measure network status and configuration parameters in order to drive its reconfiguration when needed.

The ability to think, learn and benefit from past experience requires interaction between cognitive agents. Cognitive networks are usually composed of a set of cognitive engines, which can reside in a protocol layer, between different layers, or be distributed among various nodes (Kliazovich, Granelli, & Fonseca, 2009). Each cognitive agent operates locally but contributes to the achievement of global goals by interacting with other cognitive agents. Indeed, the efficiency of the operation of cognitive networks depends on the communication among these agents. Depending on the communication scope, which can be inter-layer, intra-layer, or network wide, different technologies can be used, each imposing specific constraints on the speed and delay of information exchange. Such constraints cannot be neglected and should be taken into account during the design of cognitive network architectures.

Cognitive networking encompasses multiple wireless technologies and is designed to deal with the complexities of network configuration, as well as the support of user applications. Cognitive networks can increase the profit of wireless service providers by reducing costs and developing new streams of revenue as a result of the provisioning of heterogeneous wireless access solutions (Zhang & Hanzo, 2010). The benefits for service providers

deriving from the adoption of cognitive networks include the following: the possibility of relying on common hardware and software platforms while coping with the evolution of radio technologies, the development of new services, the minimization of infrastructure upgrades, accelerated innovation, and the maximization of return-on-investment (Clark at el., 2003).

Cognitive technologies allow network operators to continuously analyze network configuration and its performance. Moreover, reconfiguration can be triggered by application requirements, policies, or billing plans. Furthermore, cognitive networking offers extended sets of operations, allowing new ways of interaction between network operators and end-users (Demestichas, Dimitrakopoulos, & Strassner, 2006).

In the following sections, we discuss a reference network architecture which can be used to enable cognition properties, with an emphasis on aspects related to interoperability and incremental deployment. We also provide an extensive survey of various techniques which can be used for cognitive information exchange; this is the key to the design and successful deployment of such cognitive optimization techniques. Finally, the chapter presents cognitive communication protocols for the transport and the link layers.

BACKGROUND

Cognitive Node Architecture

Designing the architecture of a node in a cognitive network requires a careful balance between performance and interoperability. Indeed, interoperability represents a stringent constraint in network research.

Cognitive node is a generic network node, such as user terminal equipment, which carries complete or partial implementation of TCP/IP protocol stack and is capable of running a cognitive process for the purpose of reconfiguration

of cognitive protocols stack parameters. Figure 1 shows the main functional blocks of a cognitive node. The design is oriented to support the requirements of interoperability with the TCP/IP protocol stack. In line with this, software modules are included in protocol layers, so that information internal to the layer can be obtained (observation) and parameter values tuned (action). Information monitored for different protocol layers is delivered to the cognitive engine implemented in a cognitive node. Quality of service requirements of running applications, such as a minimum bandwidth, delay bounds and packet loss rate are provided at the application layer. Information on round trip times and achievable sending rate is available at the transport layer. In case of a mismatch between these values and those required by the applications, the cognitive node can optimize transport layer parameters to achieve higher throughput or reduce delays. The cognitive node can, for example, force the application behavior to adapt itself to the current network conditions by selecting a lower bitrate codec. The network layer provides information on the network state, routing decisions, and the

selected Network Interface Card (NIC). The link layer signals changes in the link state (e.g., fading), reports bit and packet error rates, and enables higher layers to handle handover procedures. The physical layer can report transmission rate values and control the power emitted by transmitters.

The *cognitive engine* analyzes data and makes decisions. The goal is to properly configure communication protocols adequately on different layers and optimize protocol stack performance. To do this, the cognitive engine interacts with the running applications, controlling all protocol layers, and maintaining databases with observations of past performance, accumulated knowledge, and targeted goals. The cognitive adaptation algorithms are structured in steps, such as observation, data analysis, decision-making, and action. The cognitive engine periodically reconfigures protocol parameter values, monitors the obtained performance, and stores performance information in the database. Observations are used to derive knowledge which, after combined with the setup parameter values employed by the cognitive engine, can lead to optimal performance.

*Figure 1. Cognitive node architecture*
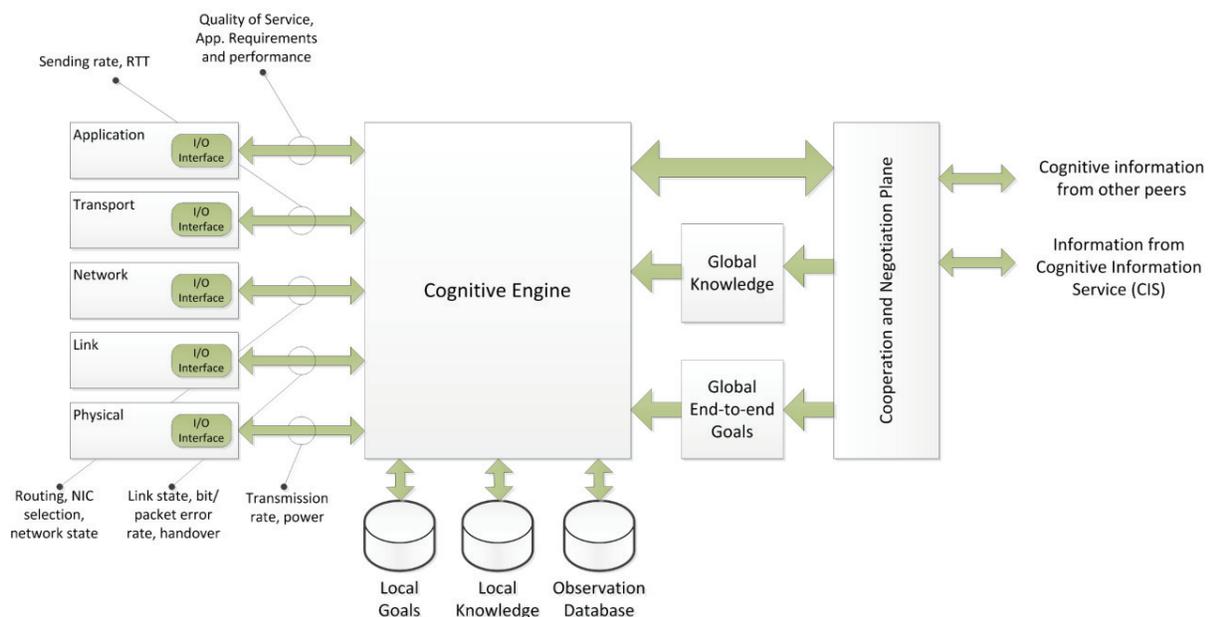
The decisions made by the cognitive engine at the node aim to optimize the protocol stack performance. They are driven by local goals for the specific nodes, specified and stored in the local database. Most goals are generated by the demands and QoS requirements of user applications. The cognitive engine also implements communication interfaces for information sharing. The decision-making process can take into account performance information available in the local knowledge base, as well that obtained from neighboring nodes.

The *cooperation and negotiation plane* attempts to perform network-wide cognitive optimization by allowing the exchange of cognitive information with other cognitive nodes in a distributed manner. Information harvesting can be either scheduled or undertaken by using instant requests or interrupts. Moreover, information can be either node-specific or data flow specific. In addition, the cognitive information can be exchanged using the Cognitive Information Service (CIS) which is a dedicated service in the network that receives and aggregates cognitive information from multiple nodes and provides information to other nodes upon request. The cooperation and negotiation plane allows nodes to obtain global knowledge, especially information on system-level goals. Nodes are required to cooperate in either a distributed or a centralized way to achieve global goals.

One of the main characteristics of the architecture of cognitive network is scalability. This is assured by the use of a combination of cognitive algorithms operating at the node level (centralized) and at the network level (distributed). Specifically, at the node level, the core cognitive techniques (such as data analysis, decision making, and learning) are concentrated on the cognitive planes of the nodes and centrally implemented. Moreover, observation and add-ons to the protocol layers are typically "non-intelligent". Distributing the cognitive process among the protocol layers (especially the learning and decision making functions) would require complex algorithms for synchronization and coordination between intra-layer cognitive processes. A simpler solution is the adoption of a centralized cognitive process.

Figure 2. Cognitive information service
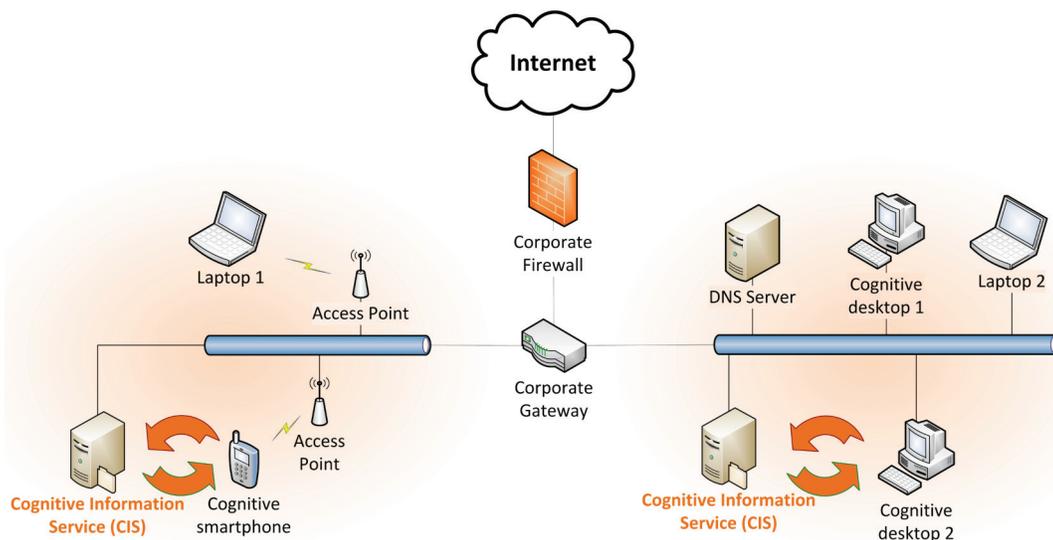
Cognitive Information Service

Cognition implies the actual or virtual existence of a database to enable the maintenance of the "history" of the network and support the decision-making process.

Figure 2 elaborates on such a concept by introducing the idea of the Cognitive Information Service (CIS) which is a network service that assists in the information exchange between cognitive nodes. It can receive, process, and aggregate cognitive information from any node in the network, as well as analyzing it and deriving system-level knowledge to guide future configuration decisions to be made by other nodes in the network.

The CIS, similar to the Domain Name Service (DNS), is implemented in a network segment with well-defined boundaries. The number of the nodes served and the size of the network segment should be limited by the design or implementation constraints to prevent CIS overload.

Both point-to-point and point-to-multipoint communications with the CIS server is possible. The latter is especially useful when the aggregation of cognitive information is sent to a large number of nodes. For this, either IP multicast at the network level or MAC broadcast at the link level can be used.

Whenever a new node joins the network, it can query the CIS for the optimal suggested configuration of its protocol stack parameters. However, the configuration offered by the CIS to a new node will not necessarily be optimal, as it will be computed without prior knowledge of the running applications, its traffic demands, and the peers with which the node is willing to communicate. However, the configuration can potentially offer a better startup performance than using fixed default values for the protocol parameters, as performed for the protocol of the TCP/IP stack.

The next presents a review and comparison between existing information signaling techniques that form the basis for implementation of different cognitive network techniques.

INFORMATION SIGNALING IN COGNITIVE NETWORKS

Many information signaling techniques are designed to overcome different limitations of the standard TCP/IP protocol reference model. Depending on the nature and requirements of information exchange information signaling solutions can be classified into in-band, on-demand, and broadcast signaling.

Signaling Methods

In-band signaling is the most effective signaling method from the point of view of overhead reduction. Cognitive information can be encapsulated into ongoing traffic flows, such as into optional packet header fields (Carrier Sense Multiple Access with Collision Detection, 1985), and delivered without waste of bandwidth resources.

Given its low overhead, in-band implementation of CIS is well suited for networks with wireless technologies used for networks access such as WiFi, WiMAX, and cellular (Zhang & Hanzo, 2010). One advantage of in-band signaling is that cognitive information can be associated with the portion of the application data with which it is delivered.

However, the main shortcoming of in-band signaling is that signaling is limited to the direction of the packet flow, making it unsuitable to cognitive schemes requiring instant communication between nodes not involved in ongoing data exchange.

Out-of-band on-demand signaling operates on a request-response basis and can be used as a complement to in-band signaling. It is designed for cases requiring instantaneous cognitive information delivery between network nodes. Since cognitive information is available at the requesting

node only after some round-trip time delay has elapsed, on-demand signaling is well suited for the Wired/Wireless LAN scenario.

One of the core signaling protocols considered in on-demand signaling is the Internet Control Message Protocol (ICMP) (Postel, 1981). Generation of ICMP messages is not constrained by any specific protocol layer and can be performed on any layer of the protocol stack. However, signaling with ICMP messages involves operation with heavy protocol headers (IP and ICMP), checksum calculation, and other procedures which increase processing overhead.

Out-of-band broadcast signaling allows low-overhead point-to-multipoint cognitive information delivery from CIS server to the network nodes located in the same segment. Broadcasting is especially suited for wireless networks involving cellular organization.

In this approach, cognitive information is encapsulated into a beacon, which is periodically broadcasted by wireless gateways (access points or base stations), thus fitting scenarios where a

delay in the delivery of cognitive information can be tolerated so that the broadcasting can be performed at regular intervals.

According to their scope, signaling techniques can also be divided into two broad categories: node-level signaling and network-level signaling (see Table 1).

Node-Level Signaling

Node-level signaling techniques provide the means for information exchange between different layers of the TCP/IP stack initially designed to be standalone and separated.

Interlayer signaling pipe was one of the first approaches used for the implementation of cross-layer signaling (Wang & Abu-Rgheff, 2003) and to allow the propagation of signaling messages layer-to-layer along the packet data flow. Signaling information, included in an optional portion of packet headers, for example into an optional portion of IPv6 header (Deering & Hinden, 1998), follows the packet processing path within the pro-

Table 1. Comparison of signaling approaches

	Signaling Method	Scope	Type of Signaling	Signaling latency	Communication overhead	In-band / Out-of-band	Direction of signaling	Packet Association
Node-level Signaling	Interlayer signaling pipe	Node	Indication	Medium	High	In-band	Path dependent	Maintained
	Direct interlayer communication	Node	Request/ Response	Low	High	Out-of-band	Path independent	Not maintained
	Central cognitive plane	Node	Indication/ Request/ Response	Low	Medium	Out-of-band	Path independent	Not maintained
Network-level Signaling	ICMP messages	Network	Indication/ Request/ Response	High	High	Out-of-band	Path independent	Not maintained
	Packet headers	Network	Indication	High	Low	In-band	Path dependent	Maintained
	Explicit Notification	Network	Indication	High	Low	In-band	Path independent	Not maintained
	WCI	Network	Indication/ Request/ Response	High	High	Out-of-band	Path independent	Not maintained
	Cross-talk	Node/ Network	Indication/ Request/ Response	High	Low	In-band	Path dependent	Maintained

tocol stack, either in a top-down or in a bottom-up manner. One important property of this signaling method is that information can be associated with a specific packet, from the protocol stack either incoming or outgoing.

Generally, the signaling method using packet structures is preferable, due to a lower processing overhead as well as flexibility, and simplicity of access or modification of encapsulated information on any protocol layer.

Direct Interlayer Communication (DIC), proposed in (Wang at el., 2003), was designed to improve Interlayer signaling by introducing of "signaling shortcuts" performed out of band. DIC allows non-neighboring layers of the protocol stack to exchange messages, skipping processing on adjacent layers. Along with reduced processing overhead, DIC avoids insertion of signaling information into packet headers, which makes it suitable for bidirectional communication.

Despite the advantages of direct communication between protocol layers and standardized signaling, the ICMP-based approach involves significant overhead for protocol processing. Moreover, it is limited to request-response actions, while more complicated signaling should be adapted to handle asynchronous events. To this aim, a mechanism employing callback functions at the node level can be employed. Such mechanism allows a given protocol layer to register a specific procedure (callback function) with another protocol layer, with execution triggered by specific event at that layer.

The Central Cognitive Plane, which is implemented parallel to the protocol stack, is probably the most widely proposed interlayer signaling scheme. Each protocol layer is extended with a tiny interface that allows the exchange of information and configuration commands to be performed. These interfaces are interconnected with a cognitive engine using a common bus.

Implementation of this signaling method could be as simple as the access to a shared database by all layers (Chen, Shah & Nahrstedt, 2003),

while more advanced implementations needs to introduce interfaces to provide access to protocol layer parameters and functions (Raisinghani & Iyer, 2006).

Network-Level Signaling

Most of the existing cross-layer signaling proposals employ signaling between different protocol stack layers of a single node. However, cognitive processes have a network-wide scope and have end-to-end goals (Thomas, Friend, DaSilva & MacKenzie, 2006; Georgakopoulos, Tsagkaris, Karvounas, Vlacheas, & Demestichas, 2012). Consequently, cognitive networks require signaling approaches capable of effectively delivering signaling information between different nodes.

Packet headers can encapsulate signaling information and propagate it along the data flow making it available to network routers as well as to the end nodes. This method of signaling keeps overhead to the minimum and allows the association of signaling information with specific network packets. However, signaling information carried in the packet header implies a fixed direction of the signaling flow.

ICMP messages is usually the default signaling method, since signaling information carried by ICMP messages can be directed and processed at the destination in the same way ordinary IP data packets are processed. Moreover, ICMP messages are usually processed in the kernel, rather than in the user application domain.

Signaling using ICMP messages is desirable when instant communication is needed. However, these messages consume network bandwidth and have an impact on the delays experienced by other flows. They should be used only as a supplement to other signaling scheme to packet headers.

Explicit notification schemes, such as Explicit Congestion Notification (ECN) (Ramakrishnan, Floyd & Black, 2001), can be employed for network-level signaling. ECN signaling is performed in-band by letting network routers mark in-transit

TCP data packets with a congestion notification bit. At the receiver, this marking is feedback in TCP acknowledgement packets directed to the sender.

The main advantage of explicit notification schemes is the low overhead; the drawbacks include the limitation of signaling propagation to a specific path, the need to maintain of the signaling loop and the requirement for all network routers to support signaling and traffic generation functionalities.

The Wireless Channel Information (WCI) adopts a Central Cognitive Plane architecture to promote network-level signaling (Kim, 2001); this is accomplished by using a specifically dedicated network service that collects current wireless channel characteristics from both the link and the physical layers. Information is aggregated and sent to the mobile stations, allowing better adaptation to constantly changing channel conditions. In this method, communication is performed using standard TCP/IP protocols, such as SNMP and HTTP over TCP.

CrossTalk is a combination of node- and network-level signaling (Winter, Schiller, Nikaein & Bonnet, 2006). CrossTalk involves two cross-layer optimization planes: one responsible for the organization of cross-layer information exchange between protocol layers of a single node, and the other responsible for network-wide coordination. Most of the signaling is performed in-band, using packet headers, making it accessible not only to the end host but also to the network routers. Cross-layer information received from the network is aggregated and then employed jointly with information on global network conditions in the optimization of local protocol stack operation.

Comparison of Signaling Approaches

In this section, a comparison of the characteristics of the different signaling type, scope, latency, communication overhead, banding, signaling

direction and association with specific packet flows in the network.

There is no general scheme optimal for coping with both node- and network-wide signaling. Therefore, various signaling methods must be employed simultaneously in cognitive networks so that efficient functionality of cognitive algorithms can be ensured.

This comparison is summarized in Table 1. Further details can be found in (Kliazovich, Granelli & Fonseca, 2011).

Scope defines the boundaries of operation. Solutions limited to a single protocol stack tend to be flexible, since they can use internal protocol stack structures such as packet format and callback functions, thus avoiding the processing related to overhead.

Solutions operating at the node are suitable for signaling between reconfigurable elements of a cognitive network. Signaling can typically be implemented using direct interlayer communication methods and either interlayer signaling pipe or a central cognitive plane.

Type of signaling is related to the communication primitives supported by each signaling method. Approaches encapsulating signaling information into packet structures, such as, packet headers on the use of explicit notification, are limited to the use of these structures, while other approaches that transmit out-of-band support a wider range of communication, including request-response actions.

The choice of signaling depends on the actions required to be performed between cognitive agents. At the node level, a cognitive engine performing blind monitoring of the environment can be connected with a cognitive engine that employs methods based on packet structures, allowing for low-overhead communication. However, when a cognitive agent needs to setup certain parameters, request-response actions are unavoidable, which implies on heavy signaling overhead.

Signaling latency provides information on the delay associated with the delivery of signaling

messages. This information is essential when signaling is to be performed across the network. The local delay resulting from signaling methods is usually several orders of magnitude less than that on the network-level. However, the interlayer signaling pipe method is slower than direct interlayer communication due to processing being carried out layer-by-layer. Moreover, the interlayer signaling pipe method can only suggest an asynchronous reaction to events, whereas direct communication allows instantaneous reactions.

Propagation delay also impacts on the efficiency of cognitive network implementation, as well as the on architecture design. It also influences information aggregation and the intervals for reporting adopted by cognitive engines involved in monitoring as well as impacting on the speed of cognitive engine, and of the reactions to changes in the network.

Communication overhead refers to the amount of network resources required for signaling. The encapsulation of signaling information into packets headers does not require additional network resources as long as reserved fields are used, leads to only a minor increase in processing when optional packet header fields are involved. The use of ICMP messages, however, requires additional overhead for delivery, which consumes network resources. The overhead for local signaling includes the number of operations (CPU cycles) required to deliver the message from one layer to another (his parameter does not include processing overhead of message encapsulation). The greatest communication overhead for local communications is associated with interlayer signaling pipe due to processing on several protocol layers prior to message delivery.

The least communication overhead is involved in signaling methods relying on existing data flow; other signaling methods, such as those using ICMP messages or WCI, require standalone transmission of signaling information encapsulated into low protocol headers, which consumes network band-

width. Moreover, it is possible that the bandwidth consumed by signaling may nullify all the benefits to be obtained from cognitive optimization if the reporting interval of the cognitive engine is not chosen appropriately.

In-band/Out-of-band parameter indicates whether signaling information is transmitted in existing data traffic flow (in-band) or is transmitted separately (out-of-band). In-band signaling does not lead to any on significant overhead in term of network bandwidth and routing resources, although the primitives which can be utilized (packet headers, explicit notification and Crosstalk) are limited and delivery latency high. On the other hand, out-of-band signaling, since it is not constrained by signaling type, allows for faster information delivery, although this is achieved at the expense of greater network resource use.

Direction of signaling is related to the applicability of a signaling approach for a cross-layer optimization scheme. Out-of-band signaling schemes are path independent and can provide fast reactions to events, moreover, these reactions can take place simultaneously. Path dependent signaling, at the other hand, involves only asynchronous reaction. The speed and flexibility of path independent signaling is achieved at the expense of additional communication resources. Moreover, it does not allow packet association.

Packet association indicates whether signaling information can be associated with a specific packet transmitted through the network. Such a property is required by many optimization approaches. For example, at the network level, an ECN signal sent with a TCP data packet and echoed back with TCP acknowledgement by the receiver indirectly carries information related to the TCP flow for which ECN signal was sent. At the node level, information monitored on the physical layer (SNR or BER) must typically be associated with the packet for which it was measured.

In-band signaling techniques maintain indirect association between transferred signaling infor-

mation transferred and the packet used to carry it. However, if out-of-band signaling is used such an association can be explicitly inserted. A good example is when the "Time Exceeded" ICMP message includes copies of the headers of the packet dropped.

Summarizing, we can observe that a large number of signaling methods exists that provide means for information exchange in cognitive network. Proper signaling methods should be selected at the design stage depending on the required scope and speed of signaling, signaling latency, and communication overhead. As a result, different cognitive networks may employ several information signaling techniques at the same time, especially those implementing cognitive process at the network level.

In the next section two examples of communication protocols implementing cognitive techniques at the transport and at the link layers are presented.

COGNITIVE COMMUNICATION PROTOCOLS

Cognitive Transmission Control Protocol

The Cognitive TCP (CogTCP) protocol (Kliazovich, Malheiros, Granelli, Madeira & Fonseca, 2010) continuously reconfigures its parameters, such as congestion window increase and decrease factors, to achieve optimal performance under constantly changing network conditions. These self-configuration capabilities are supported by a three step feedback loop: data analysis, decision making, and action.

A default value P_{def} is assigned to the parameter of interest for each a new flow. The cognitive engine periodically collects performance information on the current value of the parameter P. The average system performance obtained by using

the current value of P constantly being updated and stored in the local knowledge base. In the following equation T_a denotes system performance obtained with the current value of the parameter of interest (P). The average value of T_a is obtained using an exponentially weighted moving average (EWMA) as follows:

$$T_a = T_a * (w) + T_m * (1 - w),$$

where T_m is the current value of a system performance value obtained from using T_a, and w (the smoothing factor) is the weight assigned to the most recent performance measurement; i.e., the new value for average system performance is obtained as a sum of previously accumulated system performance and current performance measurement weighted by w.

During runtime, the values of P obtained considering a probability distribution which gives the deviation from the mean value. The smoothing factor s can be used to control the relevance of recent performance information, thus, providing a balance between stability and reactiveness to networks changes. The values of w close to zero increase the effect of the most recent performance measurements on average performance, thus enabling the algorithm to react quickly to changes while low values lead to more conservative behavior, resulting in more stable systems, since irrelevant transient changes do not affect the convergence to the optimal operational point.

During the decision making step, the cognitive mechanism continuously adjusts the mean of the distribution, in order to converge to the optimal value of P under current operational conditions. The mechanism can adjust the mean of the distribution by utilizing cognitive information and configuration policies from remote nodes or from a CIS service. Indeed, the decision making process is decentralized and can be performed by each

node independently without communication with any other node or server.

Reconfiguration requires a generation of an update value for the controlled protocol parameter. In each time iteration, a new random value is generated according to a normal distribution, and assigned to the parameter P. The standard deviation (σ) of the normal distribution defines the aggressiveness of the mechanism. The lower the value of σ, the more conservative is the behavior of the algorithm on trying new values for P. Therefore, this parameter directly affects convergence and system stability.

In TCP, controlling the evolution of TCP window allows for the adjustment of network utilization, protocol fairness, and the level of network congestion. Higher values are desirable in high bandwidth-delay network with low or moderate congestion levels, but should otherwise be avoided. However, there is no effective way for a network node to determine, in advance, the available network bandwidth and the level of congestion along the end-to-end paths between the sender and the receiver (Croce, Mellia & Leonardi, 2010).

Figure 3 presents CogTCP performance results for a setup where the information sender and receiver are interconnected by links characterized with different error rates and round trip times. CogTCP is able to keep performance close to the optimal and outperforms configurations with fixed values (represented in Figure 3 with bars 1-4) of the congestion window increase factor. These results demonstrate the benefits of dynamic adaptation in avoiding performance degradation under varying network conditions. A detailed benchmarking of the CogTCP solution is available in (Malheiros, Kliazovich, Granelli, Madeira, & Fonseca, forthcoming).

Cognitive Medium Access Control (MAC) Protocol

The Cognitive MAC (CogMAC) protocol (Kliazovich, Lima, Fonseca, Granelli & Madeira, 2009) is a technique for cognitive optimization of multiple link layer parameters, such as retry limit and contention window, in dynamic network environments. It does not rely on the explicit knowledge of network characteristics. CogMAC is implemented as a cognitive plane in parallel with the protocol stack but interfaces with all layers for the exchange of data and control information. CogMAC continuously infers dependencies between the achieved performance and protocol configuration parameters with this information, it builds and maintains a local knowledge base allowing it to trigger decisions using previous performance experience in a cognitive manner.

The data analysis, decision making, and action phases of the CogMAC protocol are similar to

Figure 3. CogTCP throughput performance
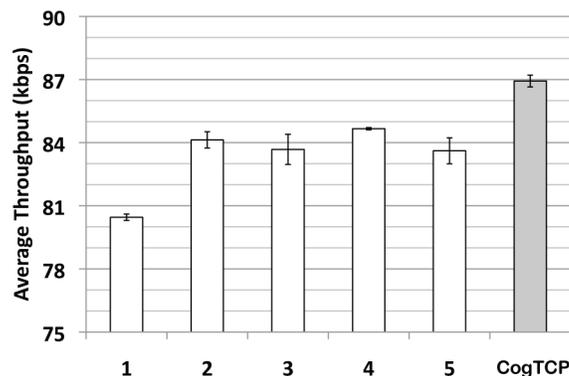
those of CogTCP, outlined in the previous section. The CogMAC performs dynamic runtime configuration of the following parameters: retry limit, contention window, and the RTS/CTS threshold when using the IEEE 802.11 link layer.

The retry limit parameter can limit the number of retransmission attempts made by a link layer before a data frame is dropped, with no notification being sent to upper layers. Low values of this parameter keep medium access delays bounded, but can only compensate for low channel error rates. High values, however, are appropriate for noisy channels although at the cost of increasing medium access delay.

The contention window parameter defines the size of the exponential backoff period and controls the rate of channel collisions. Large contention windows are recommended for dense networks where a large number of collisions can occur, while low congestion window values should be used in sparse networks to avoid unnecessary increases in medium access delay.

Finally, the RTS/CTS threshold parameter defines the minimum size of the data frame which transmission is accompanied by the RTC/CTS exchange. The optimal value depends on the number of cancelled nodes and the intensity of the traffic flow in the network.

The CogMAC approach aims to maximize the link layer throughput while ensuring bounds to the medium access delay experienced by individual nodes. Such a goal is easy to achieve when the exact number of mobile nodes is known, as well as their traffic demands and hidden nodes, and the error rates of the wireless channels. Figure 4 presents CogMAC throughput and delay performance. As the number of network nodes increases, the CogMAC shifts the points of operation to deliver optimal throughput while keeping the delay bounded. A detailed benchmarking of CogMAC solution is available in (Lima, Kliazovich, Granelli, Madeira, & Fonseca, 2013).

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This chapter presents an overview of the state-of-the-art in cognitive networking. Following the discussion on fundamental techniques of cognitive processes, reference architectures for cognitive node and cognitive network service are presented.

At the scope of a network node, the cognitive engine and the cooperation and negotiation plane, implemented in parallel to the protocol stack, control protocol configuration and performance in real

Figure 4. CogMAC a) delay and b) throughput performance

a)

b)
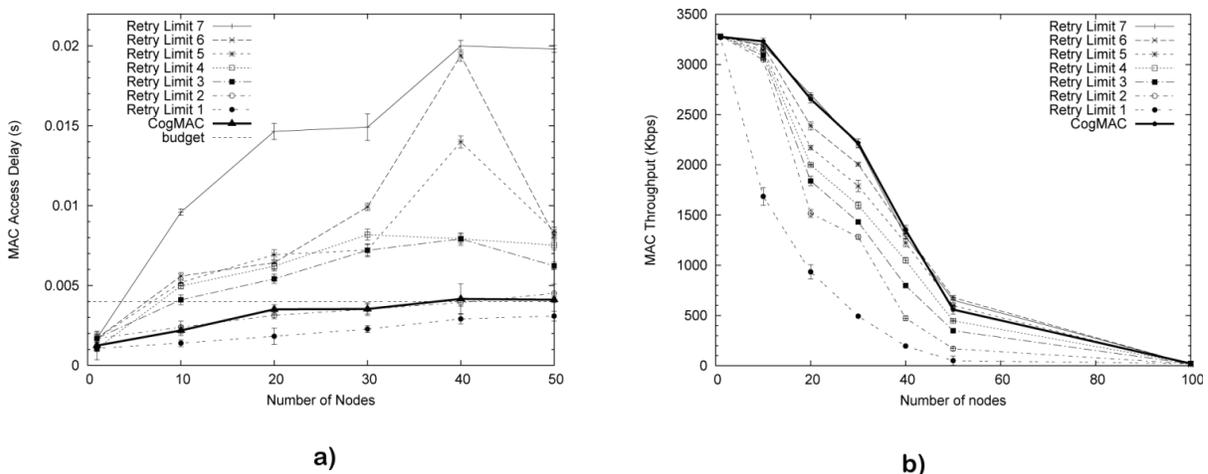
time. At the scope of the network, the Cognitive Information Service (CIS) can be implemented to foster cognitive information exchange among the network nodes, thus enabling collaborative decision making. By the scope of operation CIS service is similar to DHCP service and can be implemented to serve network segments in both wired and wireless networks.

The chapter has reviewed several node and network level signaling techniques and provides a comparison of information signaling techniques considering their scope, type of signaling, signaling latency, communication overhead, method of signaling, direction of signaling, and available information association methods. Such a comparison contributes to the discussion of the applicability of various information signaling techniques in cognitive networks.

Finally, two examples of cognitive communication protocols are presented: the Cognitive TCP which works at the transport layer and the Cognitive MAC which operates at the link layer.

Cognitive functionalities are expected to become an essential component in the design of future communication systems, networks, and protocols simplifying management and enabling adaptability. The standardization activities will focus mostly on system architectures and information signaling techniques to ensure interoperability in cognitive systems.

REFERENCES

Chen, K., Shah, S. H., & Nahrstedt, K. (2003). Cross-layer design for data accessibility in mobile ad hoc networks. *Wireless Personal Communications*, *21*, 49–75. doi:10.1023/A:1015509521662.

Clark, D. D., Partridge, C., Ramming, J. C., & Wroclawski, J. T. (2003). *A knowledge plane for the internet*. Paper presented at ACM SIGCOMM. Karlsruhe, Germany.

Croce, D., Mellia, M., & Leonardi, E. (2010). The quest for bandwidth estimation techniques for large-scale distributed systems. *SIGMETRICS Performance Evaluation Review*, *37*(3), 20–25. doi:10.1145/1710115.1710120.

Deering, S., & Hinden, R. (1998). Internet protocol, version 6 (IPv6) specification. *RFC 2460*.

Demestichas, P., Dimitrakopoulos, G., & Strassner, J. (2006). Introducing reconfigurability and cognitive networks concepts in the wireless world. *IEEE Vehicular Technology Magazine*, *1*(2), 32–39. doi:10.1109/MVT.2006.283572.

Georgakopoulos, A., Tsagkaris, K., Karvounas, D., Vlacheas, P., & Demestichas, P. (2012). Cognitive networks for future internet: Status and emerging challenges. *IEEE Vehicular Technology Magazine*, *7*(3), 48–56. doi:10.1109/MVT.2012.2204548.

Kliazovich, D., Granelli, F., & Fonseca, N. L. S. (2009). Architectures and cross-layer design for cognitive networks. In *Handbook on Sensor Networks* (pp. 2-24). World Scientific Publishing Co, Inc.

Kliazovich, D., Granelli, F., & Fonseca, N. L. S. (2011). Survey on signaling techniques for cognitive networks. In *Proceedings of the IEEE International Workshop on Computer-Aided Modeling Analysis and Design of Communication Links and Networks (CAMAD)*. Kyoto, Japan: IEEE.

Kliazovich, D., Granelli, F., Fonseca, N. L. S., & Piesiewicz, R. (2009). *Cognitive information service: Basic principles and implementation of a cognitive inter-node protocol optimization scheme*. Paper presented at the IEEE Global Communications Conference (GLOBECOM). Honolulu, HI.

Kliazovich, D., Lima, J., Fonseca, N. L. S., Granelli, F., & Madeira, E. (2009). Cognitive link layer for wireless local area networks. In *Proceedings of the IEEE Latin-American Conference on Communications (LATINCOM)*. IEEE.

Kliazovich, D., Malheiros, N., Granelli, F., Madeira, E., & Fonseca, N. L. S. (2010). CogProt: A framework for cognitive configuration and optimization of communication protocols. In *Proceedings of the 2nd International Conference on Mobile Lightweight Wireless Systems (MOBILIGHT)*. Barcelona, Spain: MOBILIGHT.

Lima, J., Kliazovich, D., Granelli, F., Madeira, E., & Fonseca, N. L. S. (2013). *CogMAC - A cognitive link layer for wireless local area networks*. ACM/Springer Journal of Wireless Networks.

Malheiros, N., Kliazovich, D., Granelli, F., Madeira, E., & Fonseca, N. L. S. (2013). *A cognitive approach for self-configuration of communication protocols*. IEEE Transactions on Network and Service Management.

Malheirosa, N., Kliazovich, D., Granelli, F., Madeira, E., & Fonseca, N. L. S. (2013). A cognitive approach for self-configuration of communication protocols. *Computer Networks*.

Maravedis Market Research and Analysis. (2006). *Clash of the titans – WiMAX and 4G: The battle for convergence is joined*. Author.

Postel, J. (1981). *Internet control message protocol*. IETF. Kim, B.-J. (2001). A network service providing wireless channel information for adaptive mobile applications: I: Proposal. In *Proceedings of the IEEE International Conference on Communications (ICC)*, (pp. 1345 – 1351). IEEE.

Raisinghani, V. T., & Iyer, S. (2006). Cross layer feedback architecture for mobile device protocol stacks. *IEEE Communications Magazine*, *44*(1), 85–92. doi:10.1109/MCOM.2006.1580937.

Ramakrishnan, K., Floyd, S., & Black, D. (2001). The addition of explicit congestion notification (ECN) to IP. *RFC 3168*.

Std, I. E. E. E. 802.3. (1985). Carrier sense multiple access with collision detection. IEEE.

Thomas, R. W., Friend, D. H., DaSilva, L. A., & MacKenzie, A. B. (2006). Cognitive networks: Adaptation and learning to achieve end-to-end performance objectives. *IEEE Communications Magazine*, *44*(12), 51–57. doi:10.1109/MCOM.2006.273099.

Wang, Q., & Abu-Rgheff, M. A. (2003). Cross-layer signaling for next-generation wireless systems. In *Proceedings of the IEEE Wireless Communications and Networking (WCNC),* (pp. 1084 – 1089). IEEE.

Winter, R., Schiller, J. H., Nikaein, N., & Bonnet, C. (2006). CrossTalk: Cross-layer decision support based on global knowledge. *IEEE Communications Magazine*, *44*(1), 93–99. doi:10.1109/MCOM.2006.1580938.

Zhang, R., & Hanzo, L. (n.d.). Wireless cellular networks. *IEEE Vehicular Technology Magazine, 5*(4), 31-39.

ADDITIONAL READING

Liu, X., Hu, S., Xiao, Y., Liu, Y., Qu, G., & Kim, K. (2010). The multiple prioritized access mechanism of cognitive networks. In *Proceedings of IET 3rd International Conference on Mobile and Multimedia Networks (ICWMNN 2010)* (pp. 168-172). IET.

Lou, J., Luo, T., & Yue, G. (2010). Power allocation for cooperative cognitive networks. In *Proceedings of the 2010 3rd IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)* (pp. 412-416). IEEE.

Mihailovic, A., Nguengang, G., Borgel, J., & Alonistioti, N. (2009). Building knowledge lifecycle and situation awareness in self-managed cognitive future internet networks. In *Proceedings of the 2009 First International Conference on Emerging Network Intelligence* (pp. 3-8). IEEE.

Nolan, K. E., & Doyle, L. E. (2007). Teamwork and collaboration in cognitive wireless networks. *IEEE Wireless Communications*, *14*(4), 22–27. doi:10.1109/MWC.2007.4300979.

Qi, J., Zhang, S., Sun, Y., & Tan, L. (2011). Research on available resource management model of cognitive networks based on intelligence agent. In *Proceedings of the 2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN)* (pp. 201-204). IEEE.

Qusay, M. (2007). *Cognitive networks: Towards self-aware networks*. New York: Wiley-Interscience.

Vu, M., Devroye, N., Sharif, M., & Tarokh, V. (2007). Scaling laws of cognitive networks. In *Proceedings of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications* (pp. 2-8). IEEE.

Wang, Z.-D., Wang, H.-Q., Feng, G.-S., Li, B.-Y., & Chen, X.-M. (2010). Cognitive networks and its layered cognitive architecture. In *Proceedings of the 2010 Fifth International Conference on Internet Computing for Science and Engineering (ICICSE)* (pp. 145-148). ICICSE.

Wentao, H., & Xinbing, W. (2011). Throughput and delay scaling of general cognitive networks, 2011. *Proceedings - IEEE INFOCOM*, 2210–2218.

Wentao, H., & Xinbing, W. (2012). Capacity scaling of general cognitive networks. *IEEE/ACM Transactions on Networking*, *20*(5), 1501–1513. doi:10.1109/TNET.2011.2180400.

KEY TERMS AND DEFINITIONS

Cognitive Information Service: A network service assisting in cognitive information exchange between network nodes.

Cognitive Network: A communication network that makes use of cognitive processes to improve operation.

Cognitive Node: A network node running cognitive process.

Cognitive Process: A process, which involves active thinking and knowledge acquisition, and uses the acquired knowledge to trigger future decisions.

Cognitive Protocol: A protocol making use of cognitive process to optimize its operation.

Information Signaling: The process of information exchange between network nodes.